

市场环境下智能配用电系统分层协同优化运行: 研究挑战、进展与展望

叶宇剑¹, 吴奕之¹, 胡健雄¹, 汤奕¹, 陈涛¹, Goran STRBAC²

- (1. 东南大学电气工程学院, 江苏省南京市 210096;
2. 伦敦帝国理工学院电气与电子工程系, 英国伦敦 SW72AZ)

Hierarchical Coordinated Optimization for Power Distribution and Consumption System Operation in a Market Environment: Challenges, Progress and Prospects

YE Yujian¹, WU Yizhi¹, HU Jianxiong¹, TANG Yi¹, CHEN Tao¹, Goran STRBAC²

- (1. School of Electrical Engineering, Southeast University, Nanjing 210096, Jiangsu Province, China;
2. Department of Electrical and Electronic Engineering, Imperial College London, London SW72AZ, UK)

ABSTRACT: With the increasing proliferation of distributed energy resources in the distribution network, how to establish an effective market-based trading mechanism in the power distribution and consumption system, while achieving efficient and coordinated optimization of market trading and power system operation has attracted unprecedented research interests in China and beyond. In the market environment, the operation and management of each layer of the power distribution and consumption system face multiple challenges, including the layer-wise increasing uncertainties, the increasing scale of market transactions, and lack of efficient coordination of market trading and safe operation of the system. This paper firstly outlines the critical scientific problems associated with optimal operation of power distribution and consumption systems in a market environment. Second, it critically reviews and summarizes existing research efforts in this area, employing conventional optimization-based solution techniques, and subsequently concludes remaining issues that deserve further research attention. Going further, this paper comprehensively reviews relevant deep reinforcement learning techniques and outlines their current applications in the examined research area, considering the primary characteristics pertaining to market trading and dispatch challenges associated with distribution and consumption system. Finally, this paper details three directions which require further research efforts,

and also dives deep in revealing how deep reinforcement learning techniques can be developed and extended to support relevant research activities.

KEY WORDS: distribution market operation; distribution system dispatch; transactive energy; demand side management; reinforcement learning; multi-agent systems

摘要: 随着分布式资源在配电网中的比例不断提高, 如何在市场化交易机制下实现配用电系统安全经济运行成为当下的研究热点。在市场环境下, 配用电系统各层的运行管理面临着不确定性逐层加剧、市场规模快速扩展、市场交易与系统安全运行难以有效衔接等多重挑战。该文首先梳理市场环境下配用电系统运行优化的关键问题; 其次, 对传统解析优化方法的研究成果与研究中仍待解决的问题进行总结; 然后, 针对配用电系统市场交易、运行优化等问题特点, 系统性介绍深度强化学习技术, 分析归纳深度强化学习在配用电系统中的研究现状。最后, 提炼出贯穿配用电系统多层多主体协同优化问题中的三重研究需求, 并对深度强化学习技术未来的应用路径与发展趋势进行展望。

关键词: 配电市场运营; 配电系统调度; 可交易能源; 需求侧管理; 强化学习; 多智能体系统

0 引言

随着“双碳”目标战略、新型电力系统、“整县光伏”等^[1-2]等重要举措相继提出, 分布式新能源(如屋顶光伏等)和新型负荷(如电动汽车等)在配电网中的占比正在急速增长, 传统电力消费者逐渐向集发、用、储能资源于一体、具有自主决策能力的电力生产者转变^[3]。2022年2月, 国家发改委与能

基金项目: 国家自然科学基金(青年基金项目) (52207082); 江苏省基础研究计划自然科学基金青年基金项目(BK20220842, BK20210243)。

Project Supported by National Natural Science Foundation of China (Young Scientific Program)(52207082); Natural Science Foundation of Jiangsu Province for Young Scientist (BK20220842, BK20210243).

源局出台了《加快建设全国统一电力市场体系的指导意见》^[4]；8月，国家能源局出台了《深化“放管服”改革优化营商环境重点任务分工方案》^[5]的通知，两者都明确提出了健全分布式发电市场化交易机制的工作要求，鼓励分布式可再生能源更加贴近用户直接交易，增强就近/就地消纳能力，并明确了在保障电网安全运行的前提下，统筹优化电力市场运行与电网调度运行的发展路径。

与此同时，国内外众多隔墙售电、能源社区、能源交易共享项目也在积极探索终端能源交易的新模式，如我国常州市宁辉分布式发电市场化交易试点项目、美国 Brooklyn 微电网项目、Olympic Peninsula 可交易能源项目、英国 Cornwall 本地能源市场项目、德国 Sonnen Community 本地能源交易项目、荷兰 Powerpeers 共享能源项目等^[6]。众多国内外实践经验引导并形成了可交易能源机制^[7-8]理论体系。该机制面向市场交易与电网调度的双重需求，旨在以价值信号为核心，融合市场和控制手段统筹管理系统内可调控资源，是充分发挥多元产消者灵活利用柔性资源提升交易的主动性、实现系统的可控性向灵活自主管理充分转化的可行路径。

可交易能源市场环境下配用电系统的运行优化问题本质为一个包含海量多元异构产消者、零售商等市场主体和配电系统运营商等系统主体的多层多主体自治协同的复杂决策问题^[7,9]。一方面，配用电系统中异构产消者分布式资源零散庞杂，受资源特性、用户类型、外界环境及决策偏好的影响而呈现出复杂多元异构特性，难以通过解析手段精准构建其物理运行模型，为各主体决策带来困难；另一方面，可交易能源市场灵活自主运营特性与配电网的安全调度管理需求难以有效衔接，具体表现为市场主体决策权的分散化和自主化、交互信息的隐私化、不同主体优化目标的多样化，与系统安全管理的全局性调控需求之间的协同困难，后者为面向产消者的市场运营及配电网调度管理的关键问题。为有效保障配网侧海量异质资源与分散主体的有序交易、共享、互动，亟待基于分层决策、自治协同的思想，探索面向新型电力系统的配电市场机制设计与运行理论方法。

本文首先对传统解析优化方法与现有研究进行综述，总结已取得的成果，分析尚存在的不足。基于以上归纳分析，进一步介绍以深度强化学习为代表的的人工智能技术基本原理与应用潜力，并结合

配用电系统优化运行核心决策问题分析其应用场景的现状与不足。最后，针对贯穿配用电系统多层多主体协同优化中尚存的研究难题，提出将深度强化学习技术向多智能体、有约束、增强感知等方面延伸的未来研究路径，提升智能体策略应对多重不确定性的能力、提升大规模异构主体策略协同优化方法的可扩展性，解决在策略优化过程中难兼顾安全约束的问题。

1 市场环境下配用电系统运行优化问题

可交易能源市场环境下配用电系统的运行优化问题主要包括3个方面：面向异构分布式资源的产消者个体能量管理策略优化、大规模产消者集群市场交易策略协同优化、配电系统运营商-产消者集群互动策略优化，如图1所示。其中，贯穿配用电系统多层多主体协同优化问题中的三重挑战可总结如下：

1) 多重不确定性层层加剧。首先，分布式资源的运行状态受自然环境、用户用能偏好等因素的影响呈动态变化，构成了产消者的内生不确定性；其次，市场中其他产消者交易行为等外部因素构成了产消者在集群协同优化中的外生不确定性；最后，大规模产消者的市场交易行为配网安全稳定运行带来了强不确定性。多重不确定性问题伴随并贯穿配用电系统各层级各主体，且逐层加剧。现有方法主要以静态优化模型为基础，解析求解最优决策方案而非可泛化策略，故在每一次场景变换和数据更新后都需要重新调用优化模型求解，缺乏利用历史优化结果进行自主学习的能力。

2) 多异构主体协难以高效协同。多类型分布式资源能量管理和大规模异构产消者市场交易策略协同优化的双重需求，形成了多决策主体、多源量测信息、多维动作空间、多重不确定性等因素交织的高维度复杂特性，且在弱(去)中心化的交易模式下，产消者的隐私保护问题不可忽视。现有多主体协同算法难以在充分计及产消者异构性与隐私保护需求的前提下，兼顾算法的可扩展性。

3) 市场化交易与系统调度运行缺乏衔接。产消者作为市场主体，基于局部信息优化分布式资源的能量管理以及与其他产消者的市场交易策略；而配电系统运营商作为系统主体，需基于全局信息对大规模产消者市场交易计划进行安全校核，再通过交易计划调整、配网可调控资源调控等市场-调度异

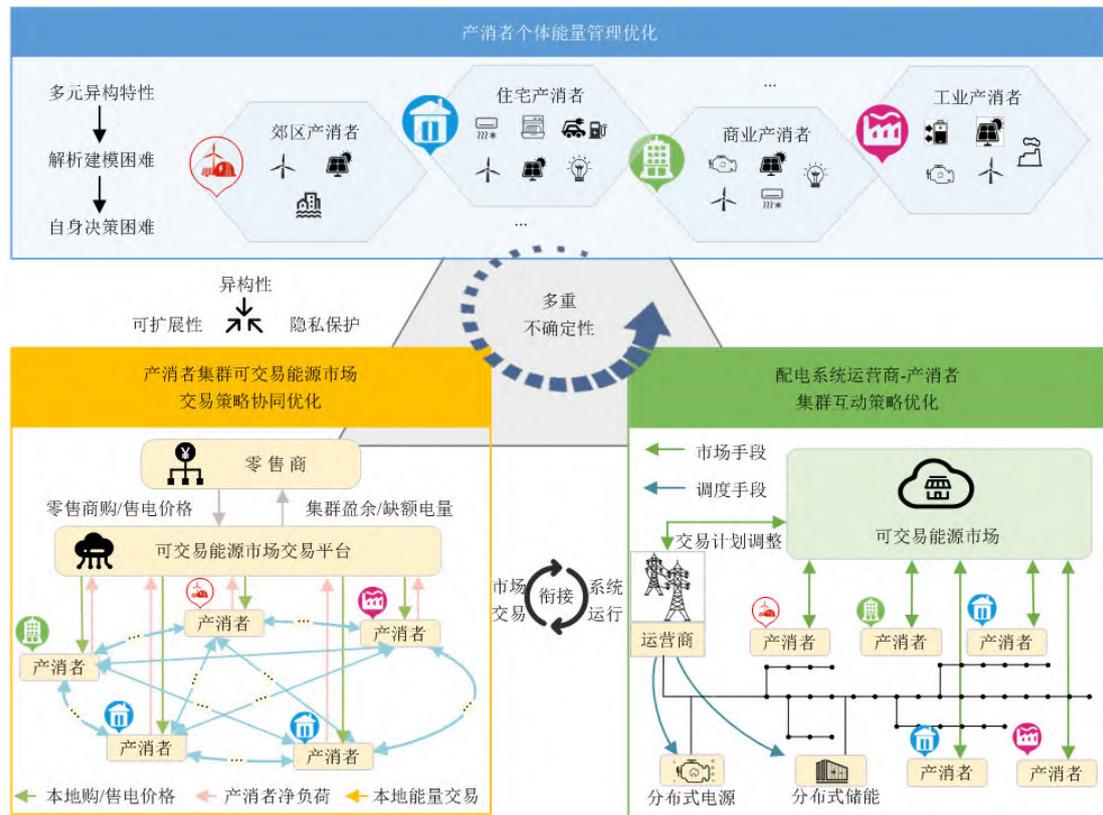


图1 可交易能源市场环境下配用电系统的运行优化问题

Fig. 1 Optimization of power distribution and consumption system operation in a transactive energy market environment

质控制手段保证其满足潮流方程等配网安全运行约束。两类主体分别从市场交易角度和物理系统运行角度进行策略优化,受信息不对称与目标不一致等因素影响,策略优化过程缺乏协同。现有研究或在市场角度偏重于优化交易策略而忽略系统约束,或在系统角度为满足系统约束而过度简化多主体交易策略协同优化过程,难以实现市场化交易与调度运行的高效衔接。

2 传统解析优化方法在配用电系统的应用

近年来,随着新一轮电力体制改革的有序推进,售电侧市场的进一步放开,基于可交易能源机制的配用电系统优化运行成为研究焦点^[7-8],呼吁通过价值信号引导异构产消者对海量柔性资源进行优化配置以促进系统优化运行的相关研究不断涌现。中国电机工程学会、IEEE 和 Elsevier 等国内外出版集团近5年中纷纷在电气工程研究领域范围内推出各类针对配电市场运营、配电系统优化运行相关的专题^[10-13],众多国内外学者均采取了借鉴运筹学领域理论的解析优化方法进行了多种技术方案尝试。针对本文所关注的配用电系统各层级的3个运行优化问题,国内外研究现状可总结如下。

2.1 产消者个体能量管理优化

目前,国内外学者针对智能楼宇、家居、园区等产消者具体形态的能量管理优化问题展开了深入研究,传统的技术路线采用基于模型最优化的思路,将负荷分为不可控负荷与可控负荷,并结合分布式电源发电特性,建立优化模型,以效用最大化为目标开展自治优化。文献[14-17]采用确定性优化模型,构建产消者分布式资源运行模型,预测负荷、光伏功率,以最小化成本为目标,优化能量管理决策。由于产消者规模较小,其DER的运行模型受外部环境、用户行为等因素的影响呈现出动态变化,DER在建模过程中涉及多种条件假设和简化,面临内生不确定性,导致其模型精度存在不足^[18-19]。此外,产消者负荷和光伏功率等内生不确定性的强随机性与间歇性,使得预测误差高于大电网的同类预测问题^[20]。以上因素可能导致分布式资源的灵活性难以得到充分挖掘,直接影响确定性优化的性能。

针对不确定性问题,文献[21-23]采用基于场景的随机规划方法,假设光伏、风电功率等不确定性参数所对应的概率分布模型,通过蒙特卡洛模拟生成对应的场景。然而,受多重外部随机因素的影响,

所假设的概率分布与实际分布偏差较大。另外，随机规划问题的求解规模随着不确定性场景数目的增加而急剧扩大，虽然场景削减技术可减少所采用场景的数量，但是该方法易导致小概率场景在削减过程中的流失，难以保证决策方案在小概率事件下的可行性。相比于随机优化，文献[24-25]采用鲁棒优化，不需要设定不确定性因素的概率分布函数，而采用不确定性集合表示其变化范围，并寻求可以应对不确定性因素的所有解^[26]。然而，鲁棒优化根据不确定性集合中的最劣场景直接进行决策，优化结果存在一定的保守性^[27]。

2.2 产消者集群可交易能源市场交易策略协同优化

根据市场协同优化方式和信息交互机制，现有产消者集群市场交易策略协同优化的研究可概括性划分为 3 类^[28]：协同型(coordinated)市场、分散型(decentralized)市场和社区型(community)市场，如图 2 所示。

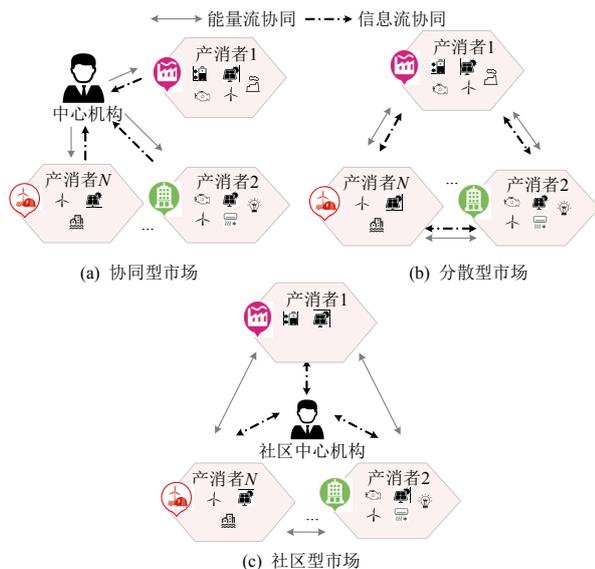


图 2 产消者集群市场交易策略协同优化研究分类
Fig. 2 Framework classification for coordination of market trading strategies of prosumer population in a transactive energy market environment

协同型市场中^[29]市场优化和信息交互均以集中式架构进行，由中心机构收集市场参与者信息并求解最优交易方案。文献[29]立足园区，提出了一种分布式能源共享服务机制，建立了以用户净用能成本最小为目标的协调运行模型。文献[30]设计了一种智能家庭间的电能交易模型，以最小化系统总成本为目标，给出了家庭光伏和储能系统的最优运行策略。文献[31]提出了一个基于社区能源库的分

布式交易模型，赋能具有需求响应能力的产消者参与可再生能源的交易与共享。文献[32-33]分别采用基于场景的随机规划与鲁棒优化模型，在协同型市场交易决策优化中考虑了产消者的负荷与光伏不确定性。协同型市场理论上可求得最优解，但在计算复杂度高导致难以扩展求解规模、隐私性保护差等问题^[34]。

分散型市场的协同优化和信息交互均基于产消者间的自主交互和优化进行。文献[35]提出了基于共识算法的分布式协调方法，通过电价信号实现产消者间的分布式电能交易。文献[36-37]提出了基于双边合同的本地交易匹配机制，产消者可根据自身偏好与其他用户协商，接受或拒绝交易。分散型市场能量管理决策和市场价格仅取决于参与者间的双边协调，在可扩展性和隐私性上具备优势，但存在无法保证市场整体社会福利最大化，收敛速度慢等局限。

社区型市场融合了集中式信息交互和分布式协同优化，设立的中心机构仅基于收集的参与者有限信息，从社区整体角度求解最优交易方案并确立合适的价值信号，引导参与者进行分布式交易。文献[38]针对能源社区中大规模产消者需求响应运行决策问题，采用交替方向乘子法制定能源社区分布式优化调度策略。文献[39]采用蒙特卡洛场景生成法进一步对分布式资源的不确定性进行建模并求解随机分布式优化调度策略。文献[40-41]基于连续双边拍卖机制搭建了日前阶段产消者电能交易框架，实现社区微网系内产消者之间能量交易与共享。社区型市场可以平衡可扩展性、隐私保护性、市场整体最优等方面的需求，在实际中应用最为广泛。此外，现有社区型市场管理多采用基于解析模型的分布式决策方法，如交替方向乘子法^[38]、协调对偶分解算法^[42]和一致性对偶分解算法^[43]等。但该类方法需要产消者与社区运营商之间频繁进行能量管理与交易信息的迭代式交互，面临大规模异构产消者内生及外生不确定性引发的计算困难，难以获得泛化性的市场交易策略。

2.3 配电系统运营商、零售商与产消者集群间的互动策略优化

大规模产消者的市场交易行为有可能引发线路阻塞等配电网安全运行问题，网损增加等配电网经济成本问题。如何实现产消者自身策略优化和系统安全管理策略优化之间的协同是亟待解决的难

题。文献[44-45]仅考虑接入在同一低压变电站的社区能源交易问题以及电网容量阈值对社区净负荷的约束,缺乏海量产消者广域配电网交易行为是否满足配网安全约束的考虑。由于个体产消者难以获取关于网络拓扑及参数等的全局信息,无法在自身决策中纳入配网安全约束,通常依靠配电系统运营商基于包含拓扑空间特性的全局信息对市场交易计划进行校核与调整。文献[46]提出了面向配电系统运营商的集中式优化模型,该模型计及了分布式资源的运行不确定性,通过最优潮流与概率分布式节点电价确定本地能量市场中产消者的交易结果。

除单层集中式优化模型外,现有配电系统运营商与产消者集群互动优化的解决方案主要采用解析双层优化模型,在上层构建考虑网络约束的配电系统运营商决策模型,在下层建立产消者集群的决策模型。文献[47-48]采用 Stackelberg 博弈(主从博弈)构建产消者与系统运营商间的互动优化模型,运营商是领导者(leader),以自身运营效益最大化为目标决策动态电价并保证网络约束的满足,产消者是跟随者(follower),以自身用电成本最低为目标,根据电价来决策自身的用电计划。但该类研究未考虑产消者之间的本地电能交易与共享行为,以及后者对配网运行状态的影响。文献[49-50]采用协同型市场机制,通过最大化产消者集群的社会福利求解产消者的交易决策;文献[51]则采用社区型市场机制,采用交替方向乘子法(alternating direction method of multipliers, ADMM)制定产消者的分布式优化交易决策;文献[49-51]均在上层问题中建立配电系统运营商的决策模型对产消者市场交易计划进行安全校核,最终通过启发式算法进行迭代求解。然而,此类优化建模手段需要根据优化理论不断拆解衔接子问题和主问题,中间计算结果需要频繁信息交换。此外,由于产消者隐私保护需求,在实际中配电系统运营商难以获得下层决策模型,该理论最优安全决策缺乏实际可行性。即使忽略隐私保护问题,大规模产消者市场交易结果的强不确定性对于配网安全管理决策也是巨大挑战,现有构建随机双层规划的方法面临严重的维度灾问题,亦不具备动态感知不确定性、生成可泛化策略的能力。

当产消者发展到一定数量后,考虑到群体特性、市场行为和商业模式,产消者资源灵活性的调用还需考虑其与零售商间的互动交易优化。文献[52]针对传统售电模式(即产消者与零售商独立进行交

易),同样采用双层优化模型进行两者的决策协同。文献[53]则在上层零售商的定价决策模型中进一步考虑了配电网网络约束与零售商向配电系统运营商支付的过网费成本。由于商业模式的多元化,类似的主从博弈的关系存在于多种配用电系统场景下,例如含多电动汽车用户的负荷聚合、多光伏-储能产消者主体的微电网、含热电联产及光伏产消者的区域综合能源系统。这些系统中电动汽车负荷聚合商^[54]、微电网运营商^[55]、综合能源服务商^[56]均可以作为领导者,产消者或用户作为跟随者,前者通过求解双层优化模型决策动态价格,以零售商收益最大化目标,形成对用户侧价格响应的引导。然而,现有以零售商为研究对象的文献较少考虑产消者间的分布式交易行为,而以产消者集群为研究对象的文献通常假设零售商无法进行价格决策优化,导致两者的互动交易策略缺乏有效协同。

3 深度强化学习在配用电系统的应用

受 Google DeepMind 公司在游戏领域取得人工智能技术巨大成功^[57]的启发,深度强化学习(deep reinforcement learning, DRL)开始受到各工程科学领域,尤其是电力系统领域学者的重视。该技术作为一种自主学习行为策略的人工智能技术,将深度神经网络的普适性函数逼近能力与强化学习的策略提升能力相结合^[58],在缺乏环境对象先验知识与未来状态准确预测信息的条件下,通过与环境的交互过程学习行为策略实现回报最大化。不同于基于解析优化模型的决策方法在环境对象变化后,需要对应调整解析模型和重新求解具体决策,基于 DRL 的策略生成方法可以通过与复杂环境交互来学习具有泛化性的普适策略,具有单次计算复杂度低、环境变化自适应、无先验知识依赖等优势^[59]。针对不同的电力系统应用背景中决策主体的数量可划分为 1) 单智能体深度强化学习(single-agent DRL, SADRL); 2) 多智能体深度强化学习(multi-agent DRL, MADRL)等两大类 DRL 算法。

3.1 单智能体 DRL 算法及其应用

SADRL 算法采用单一智能体,通过马尔可夫决策过程(Markov decision processes, MDP)与应用环境进行交互,通过不断试错积累训练数据,并依据应用环境特点选择合适的 DRL 算法优化智能体策略,从环境交互中最大化获得的累计奖励^[60-61]。由于采用单一智能体感知当前(或前若干时刻)环境

状态进行决策，该方法适用于优化调度等集中式序贯决策场景^[62-63]。

3.1.1 马尔可夫决策过程

如图 3 所示，MDP 作为 SADRL 的数学基础，通常由状态空间 S 、动作空间 A 、状态转移概率函数 P 、奖励函数 R 以及智能体策略 π 等元素构成；该 MDP 结构应用于配用电系统个体主体典型优化问题时各元素的定义如图 3 中表格所示。

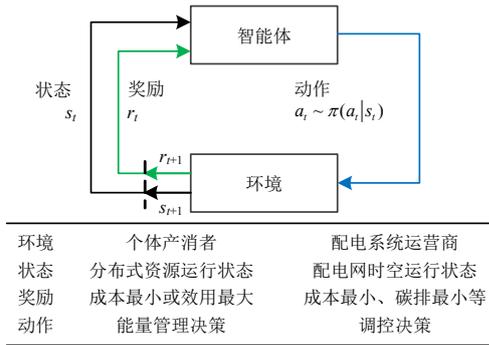


图 3 马尔可夫决策过程及其在配用电系统个体主体典型优化问题中的应用

Fig. 3 MDP and its application in representative single decision-maker optimization problems in power distribution and consumption systems

在 MDP 过程中，任意 t 时刻的智能体与环境的交互可概括为以下步骤：

- 1) 智能体依据当前环境状态 s_t 从当前策略 $\pi(s_t|a_t)$ 中采样动作 a_t 进行决策；
- 2) 环境依据状态转移概率函数 $P(s_{t+1}|s_t, a_t)$ 转移到下一时刻系统状态 s_{t+1} ；
- 3) 奖励函数 $R(s_t, a_t, s_{t+1})$ 依据智能体决策与环境状态转移情况计算当前步奖励 r_t ；
- 4) 完成上述 3 个步骤后，进入 $t+1$ 时刻的决策过程。

在智能体参与的 MDP 过程中，系统下一时刻 ($t+1$ 时刻) 系统状态的转移结果由当前时刻系统状态和智能体决策结果共同决定，整个过程中仍保持了马尔可夫性质。

基于上述 MDP 过程，DRL 算法旨在依据智能体策略与环境交互过程中记录的大量 $\langle s_t, a_t, r_t, s_{t+1} \rangle$ 四元组样本，通过最大化总决策步长 T 内的累计奖励期望 $J(\pi)$ 实现策略 π 的优化：

$$\max_{\pi} J(\pi) = \mathbb{E}_{\tau \sim \pi} \left[\sum_{t=0}^T \gamma r_t \right] \quad (1)$$

式中： $\tau = \{ \langle s_t, a_t, r_t, s_{t+1} \rangle \}$ ，表示 MDP 过程中智能体与环境交互所产生的四元组集合； T 为决策时段总

数； $\gamma \in [0, 1]$ 为折扣因子用于平衡短期奖励与长期奖励， γ 越趋近于 1，则策略越重视长期奖励。

3.1.2 单智能体 DRL 算法

随着 SADRL 技术的发展，涌现了一批适用于求解具有不同特征 MDP 问题的强化学习算法。这些算法依据待求解问题动作空间的离散或连续性划分为基于值估计的算法(value-based)与基于策略提升的类算法(policy-based)；也可以依据是否仅由当前在线交互策略产生的四元组样本进行训练划分为在线策略算法(on-policy)与离线策略算法(off-policy)；还可以依据策略输出为实际动作或动作分布划分为随机策略(stochastic policy)类算法与确定性策略(deterministic policy)类算法。表 1 展示了目前主流的 SADRL 算法的分类情况。

表 1 现有典型 SADRL 算法分类

策略	算法	随机策略	确定性策略
在线策略	policy-based	AC, A2C, A3C, TRPO, PPO	—
离线策略	value-based	—	DQN, Double DQN, Dueling DQN, Double Dueling DQN
	policy-based	SAC	DDPG, TD3

1) 基于值估计的 DRL 算法。

以深度 Q 网络(deep Q-network, DQN)及其衍生的 Double DQN, Dueling DQN, Double Dueling DQN 算法为代表的基于值估计的 DRL 算法，在经典强化学习算法 Q-learning 的基础上，通过深度神经网络拟合动作价值函数 Q，评估在当前时刻 t 系统状态 s_t 下采取各个动作 a 对应的 Q 值，从而通过选择对应最大 Q 的动作 a_t 作为决策结果输出。由于该算法需要比较各个可能动作对应的 Q 值，虽然其在离散动作空间内具有较好的性能，但难以适用于连续动作空间。

2) 基于策略提升的 DRL 算法。

为实现连续动作空间上的自主决策，出现了一类以演员-评论家(actor-critic, AC)结构为代表的 DRL 策略提升算法。该类算法通过结合值函数估计方法与策略梯度(policy gradient, PG)方法，构建用于生成决策动作的策略网络(actor 网络)，构建价值估计网络(critic 网络)，以评价策略网络性能以指导策略网络优化。一方面通过策略网络直接生成动作，克服了动作价值比较导致的离散动限制，另一方面通过值函数的评价，指导策略网络更新，避免

了PG算法对重复采样完整轨迹的依赖。这一类算法包括采用动作状态价值网络的深度确定性策略梯度(deep deterministic PG, DDPG)算法与双延迟深度确定性策略梯度(twin delayed deep deterministic PG, TD3)算法,采用状态价值网络的ACs算法(包括AC及其衍生的优势动作评(advantage AC, A2C)、异步优势动作评价(asynchronous advantage AC, A3C)算法),信任区域策略优化(trust region policy optimization, TRPO)算法与近端策略优化(proximal policy optimization, PPO)算法,以及两者兼用的柔性动作评价(soft AC, SAC)算法等。

3) 在线策略算法与离线策略算法。

在线策略算法的四元组样本生成与训练使用同一组策略和价值评价网络。由于智能体采用当前策略,在线交互产生样本,对策略进行更新,因此训练样本具有更小的方差,有利于策略提升。这一类算法包括ACs算法,TRPO算法与PPO算法等。由于在线策略算法在线采集的四元组样本仅用于本次算法更新,样本利用率低,更适用于低交互计算成本的问题,或通过采用将算法改进为分布式并行计算以提升性能(如分布式PPO与A3C算法)。

与在线策略算法相对,以DQN算法、DDPG算法、SAC算法为代表的离线策略算法引入了经验池,用以存储不同更新阶段策略交互产生的四元组样本,并通过从经验池中小批次采样,对策略进行训练更新。因此,离线策略可充分利用交互的历史数据,相比于在线算法更适用于具有高交互计算成本的问题,且由于策略实际通过离线经验池进行更新,可以通过在经验池中引入先验知识样本数据,提升策略性能。

对于离线策略类算法,由于训练样本来源于不同策略的交互过程,不可避免地存在训练样本方差过大导致的难以收敛问题。为缓解这一弊端,TD3算法在DDPG算法基础上,参考Double DQN算法引入额外的Critic网络,通过双Critic网络有效降低了样本估计的方差,从而提升了算法的收敛性能。

4) 随机策略算法与确定性策略算法。

随机策略输出连续动作的分布参数,再通过依据输出分布的随机采样进行决策,因此其在决策过程中实现了对动作空间的探索,这类算法包括ACs、TRPO、PPO、SAC等。而确定性策略算法的策略网络直接输出动作值,需要人为地引入噪声进行探索,增加了调参难度,该类算法包括DQNs、

DDPG和TD3等。

两类算法在数据利用率方面,与其探索效率相反,具有更高探索效率的随机策略算法需要训练数据和更多训练计算成本用于收敛,而确定性策略在牺牲探索效率的同时,可以更高效地利用训练数据,具有更低的训练计算成本与训练样本需求。因此,依据两类算法特点,随机策略算法更适用于具有高维动作空间且交互计算成本小的场景,确定性策略更适用于交互计算成本高且动作空间维度低的场景。为缓解两类算法固有缺陷,在高交互计算成本的场景下,随机策略算法可以通过分布式并行计算提升样本产生与训练效率。而对于确定性策略算法,考虑到大多数确定性策略算法同时属于离线策略算法,因此在具有高维动作空间的场景下,可以通过在经验池中引入先验知识,使得策略更易在最优解附近探索,弥补探索能力的不足。

3.1.3 单智能体DRL算法在配用电系统中应用

SADRL算法由于采用单一智能体进行集中感知与控制,在配用电系统中主要应用于单一主体应用场景,主要包含面向系统主体的配电网运行调度、面向市场主体的微网(用户)能量管理、零售商/负荷聚合商与用户间的互动交易等领域。

1) 配电网运行调度。

在配网的运行调度场景下,SADRL算法常用于训练调度决策智能体,通过集中式调控管理配网中各类如新能源,储能,电动汽车等分布式资源,或采用价格等经济激励对分布式可调控资源进行分散式调控,以降低配网运行成本,实现配网的绿色经济安全运行。文献[64]针对无功调节设备控制问题,采用AC结构的A2C算法,以最小化网损与动作成本为优化目标,训练智能体自主控制配网中离散无功调节设备的投切指令,实现了配网在线无功优化。文献[65]则采用适用于离散动作空间决策的DQN算法将分布式电源(distributed generator, DG)逆变器调节比例纳入动作空间。在此基础上,文献[66]针对配网中分布式电源的连续无功调节能力,采用DQN算法与AC算法分别控制长周期无功调节装置的离散投切指令与短周期DG的连续无功输出,实现了配网多时间尺度的在线无功优化。文献[67]充分考虑了配网用户的需求响应能力,采用Dueling DQN算法控制可中断负荷,通过尽可能少的负荷响应动作,最大程度降低配网运行成本。文献[68]则进一步计及微电网调控能力,采用

DDPG 算法调节微网与配网交换功率，最小化配网节点电压偏差与网损，实现了配网的安全经济运行。面向基于价格的配网分散式调控，文献[48]在主参数电价之外引入辅助参数以缓解柔性负荷在低电价时段大量聚集产生新峰值的过响应问题，采用 TD3 算法在保护用户隐私的同时优化辅助参数，降低了配网运行成本与网损、缓解了输电阻塞。

除正常的运行调度场景，SADRL 技术还被应用于气象灾害等导致配网故障后的重构与恢复调度优化中。文献[69]将配电网关键负荷恢复问题建模为 MDP，从而采用 DQN 算法控制配网分段开关，调控微网盈余电力供给关键负荷。文献[70]考虑到配网的拓扑结构特点，将图神经网络技术与 DQN 算法结合，采用图强化学习实现配电网的顺序恢复。文献[71]考虑配网中 DG 在灾后的支撑能力，将配网遭受极端事件后的 DG 调度问题建立为无转移概率的 MDP，采用 DDPG 算法进行求解。文献[72]则将灾后配网重构问题转化为双层动态优化模型，采用 PPO 算法优化决策外层的离散联络线投切指令，采用传统优化方法求解 DG 出力与负荷恢复量，并计算外层决策奖励，从而实现配网的快速恢复。

2) 微网(用户)能量管理。

在配用电系统中，微网能量管理场景与配网运行调度场景类似，但调控对象由配网中各类无功调节设备、分段开关、DG 转为各类可调节负荷与分布式储能装置等微网可控资源。文献[73]计及微网中储能、DG、可控负荷等多种可控分布式资源，以最小化配网运行成本为目标，通过将能量控制动作离散化，采用 DQN 实现了微网的能量管理，文献[74]则进一步分析对比了不同超参数设置下的算法收敛性。文献[75]则通过引入多参数动作探索机制改进 Dueling DQN 算法，以控制微网中 DG、储能与可控负荷元件，实现微网自主经济运行的同时，相比于传统 DQN 与 Dueling DQN 算法，显著的提升了性能。文献[76]则针对家庭能量管理系统，将电动汽车充电动功率纳入动作空间，并在能量管理的 DMP 设计过程中计及需求响应影响，采用集成 DQN、Double DQN、Dueling DQN 的 Rainbow 算法，提出了运行成本更低的微网能量管理策略。然而，实际的微网控制动作如 DG 功率调整、储能充放电功率调整大多为连续动作，采用基于值估计的 DQNs 算法进行的动作离散化操作，不可避免地会带来网格误差。文献[77]针对家庭能量管理问题，

采用基于策略优化的 DDPG 算法实现了对储能、空凋功率的连续控制。文献[78]充分考虑了分布式新能源出力、用能需求与价格的不确定性，通过随机策略的 PPO 算法实现了相比于 DDPG 算法 6% 的运行成本缩减。文献[79]则在 PPO 算法基础上，分别使用伯努利分布与高斯分布拟合离散、连续动作分布，并引入长短期记忆(long short-term memory, LSTM)网络提取状态时序特征，提出了高效应对多重不确定性的连续-离散混合能量管理策略。

3) 零售商/负荷聚合商-用户间的互动交易。

在零售商/负荷聚合商与用户之间的互动交易场景下，由于市场成员数量较多，因此通常采用多领导者-多跟随者斯塔克尔伯格博弈(Strackleberg game)对零售市场中的市场成员进行建模。然而，零售商因用户隐私保护而无法获得分布式资源详细运行模型与参数，难以通过解析方法求解双层优化问题获得最优定价决策。此外，受用户用能特性和决策偏好影响，其决策模型伴随着强不确定性。传统通过构建随机双层规划模型虽能在零售商定价决策中一定程度上考虑该不确定性，但面临着严重的维度灾问题，亦不具备生成可泛化性定价策略的能力。最后，传统双层优化模型的解析求解方法需假设下层优化问题为凸优化，而在刻画用户侧异构分布式资源的运行特性时，往往难以通过纯连续变量与约束实现(如可平移但不可中断负荷的运行特性)。因此，传统方法难以计及此类非凸运行特性对零售商定价决策的影响。

SADRL 的引入可有效解决上述难题，文献[80-81]设计了零售商智能体，将各用户柔性负荷的价格响应问题封装为智能体所交互的环境，进而提出了基于 Q-learning 算法的零售商动态定价策略，在不依赖环境详细信息和先验知识、不侵犯用户隐私性的条件下，学习对用户侧不确定性具有泛化性的定价策略。文献[82]进一步考虑了零售商参与电力批发商的竞价与购电问题，提出基于 DDPG 算法的竞价-定价联合策略，提升了在应对批发市场价格与终端电力用户多维不确定性时的鲁棒性。文献[83]站在电动汽车聚合商视角，在电动汽车的价格响应中考虑了离散的充/放电等级，提出了基于优先深度确定性策略梯度(prioritized DDPG, PDDPG)的聚合商定价策略，PDDPG 算法在 DDPG 的基础上引入了优先经验回放(prioritized experience replay, PER)机制，采用重要性采样方法将经验按

重要性排序,增大重要经验被采样到的概率,进一步增加了策略质量与学习效率。

3.2 多智能体 DRL 及其应用

MADRL 算法采用两个以上的智能体,依据场景特点选择合适的 MADRL 训练框架,以组织不同智能体与环境间、不同智能体间通过马尔可夫博弈模型(Markov game, MG)进行交互。MADRL 算法则依据各智能体交互过程积累训练数据,集中式或分布式地优化各智能体策略以其累计奖励^[84-85]。由于 MADRL 中各智能体决策相对独立的同时相互配合,在配用电系统中,更适用于能量交易等分布式序贯决策场景^[86-87]。

3.2.1 马尔可夫博弈过程

如图 4 所示, MG 作为 MADRL 的数学基础,由与 MDP 模型类似状态空间、动作空间、状态转移概率函数、奖励函数以及智能体策略等要素构成;该 MG 结构应用于配用电系统主体集群典型优化问题时各元素的定义如图 4 中表格所示。

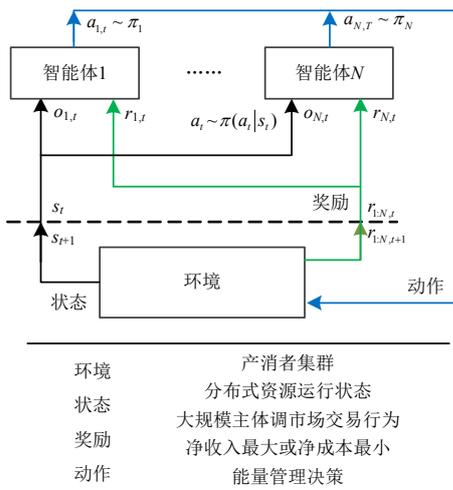


图 4 马尔可夫博弈及其在配用电系统主体集群典型优化问题中的应用

Fig. 4 MG and its application in representative multiple decision-makers optimization problems in power distribution and consumption systems

在 MG 中,其动作空间 $\{A_{1:N}\}$ 、奖励函数 $\{R_{1:N}\}$ 均为 N 个智能体 $\pi_{1:N}$ 的集合;状态空间 S 与状态转移概率函数 P 为各智能体共享,但在某些场景下,由于量测、通信、隐私等因素的限制,智能体 i 只能感知全部状态空间中的部分观测空间 $o_{i,t}$,并以此进行自主决策,此时 MG 模型即为部分可观的马尔可夫博弈(partial observable MG, POMG)模型:MG 模型中各智能体由于量测、通信、隐私等因素的限制只能感知部分观测空间,而非完整的状态空间进

行协同决策,因此 MG 模型具有部分可观性。例如,在配用电系统中多产消者能量交易策略协同优化的场景中,所有产消者的设备运行状态以及能量交易信息共同构成了全体状态空间,而各产消者只能直接量测到自身所属设备的运行状态,以及通过社区交易平台获取的其他产消者能量交易信息,无法获取如其他产消者本地设备运行状态的信息。

在 MG 模型中,任意 t 时刻的各智能体与环境的交互可概括为以下步骤:

- 1) 智能体 i 基于当前全局状态 s_t (或从获取的部分观测量 $o_{i,t}$),从当前自身策略 $\pi_i(a_{i,t}|o_{i,t})$ 中采样动作 $a_{i,t}$ 进行决策;
- 2) 环境依据状态转移概率函数 $P(s|s_t, a_{1:N,t})$ 转移到下一时刻系统状态 s_{t+1} ;
- 3) 环境采用奖励函数 $R_i(s_t, a_{1:N,t}, s_{t+1})$ 依据各智能体决策情况与环境状态转移情况计算各智能体当前步奖励 $r_{i,t}$;
- 4) 完成上述 3 个步骤后,进入 $t+1$ 时刻的决策过程。

基于上述过程, MADRL 算法求解 MG 模型的目的在于,依据 MG 交互过程中存储的大量样本,通过最大化总决策步长 T 内的各智能体累计奖励期望 $J(\pi_i)$,实现各智能体策略 π_i 的优化:

$$\max_{\pi_i} J(\pi_i) = \mathbb{E}_{\tau_i \sim \pi_i} [\sum_{t=0}^T \gamma r_{i,t}] \quad (2)$$

式中 $\tau_i = \{s_t, a_{i,t}, r_{i,t}, s_{t+1}\}$ 表示交互过程中智能体 i 与环境交互所产生的四元组集合,当采用 POMG 模型时为 $\tau_i = \{o_{i,t}, a_{i,t}, r_{i,t}, o_{i,t+1}\}$ 。

3.2.2 多智能体 DRL 框架

在 MADRL 场景中,各智能体策略执行的效果不仅依赖于场景本身特性,还会受到其他智能体决策的影响。因此,是否进行智能体间的数据信息共享将直接影响智能体策略的训练与执行效果。

依据智能体训练与执行时的集中式(智能体间进行数据信息的交互)与分布式(智能体间不进行数据信息的交互)差异, MADRL 算法框架可以划分为分布式训练-分布式执行(decentralized training decentralized execution, DTDE)、集中式训练-集中式执行(centralized training centralized execution, CTCE)和集中式训练-分布式执行(centralized training decentralized execution, CTDE)及其所衍生出来的参数共享与平均场近似两种变种算法等共 5

类，如图 5 所示，其算法框架特点如表 2 所示。

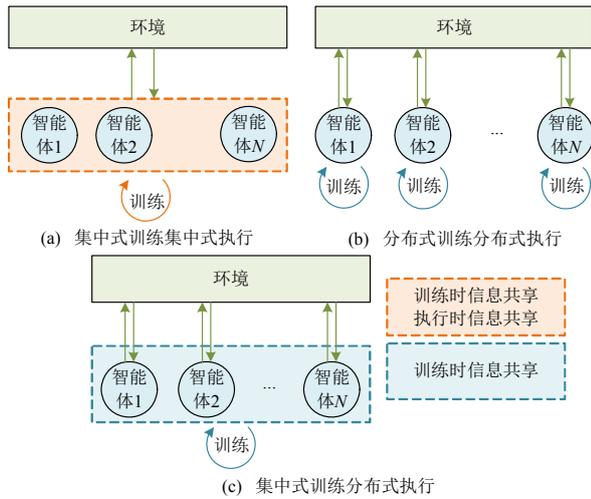


图 5 现有 MADRL 方法框架示意图

Fig. 5 Illustration of existing MADRL frameworks

表 2 现有 MADRL 方法框架特征

Table 2 Features of existing MADRL frameworks

特点	DTDE	CTCE	CTDE	CTDE 参数共享	CTDE 平均场近似
训练	分布式	集中式	集中式	集中式	集中式
执行	分布式	集中式	分散式	分散式	分散式
非静态性	是	否	否	否	否
隐私性	是	否	是	是	是
可扩展性	否	否	否	是	是

1) DTDE。

在 DTDE 框架下，各智能体间相互独立，在训练与应用时仅各自与环境进行交互，而不进行任何相互间的数据交互。由于各智能体相互独立，因此对于任意智能体可以将其余智能体视为交互环境的组成部分，从而可以直接使用 SADRL 算法对各智能体分别进行训练。

DTDE 框架的提出最早是为了应对可扩展性——随着场景维度的增加，单一智能体将遭遇维度灾的挑战。通过将高维场景的控制决策问题进行合理规划，可以很方便地使用 SADRL 算法训练多个单智能体共同交互决策。

虽然 DTDE 框架采用相互独立的多个智能体解决了场景的可扩展性问题，但这也会导致环境的非静态性，即由于各智能体将其余智能体作为环境组成部分，随着智能体策略的不断更新，各智能体交互的“环境”也将动态变化，将导致各智能体策略的收敛困难。

2) CTCE。

CTCE 框架与 DTDE 框架正好相反，在 CTCE 框架下，各智能体在训练与应用时进行完全的信息

交互，即各智能体均能感知到其他智能体的感知状态与决策结果，从而消除了环境的非静态性，并可以使用一个集中式的执行器学习所有智能体的联合策略分布。

虽然 CTCE 框架可以克服 DTDE 框架下的环境非静态性问题，可以实现智能体间的协同决策，但随着智能体数量的增加，需要进行共享的各智能体状态-动作维度将爆炸性增长，致使其缺乏对智能体数量的可扩展性。

为克服这一弊端，可采取一些改进的措施，如将集中各智能体的联合动作估计拆分为对各个智能体进行单独估计，或对集中式的值估计采用近似拆分方法，但从整体上看 CTCE 采用近似 SADRL 的模式难以完全克服环境与智能体数量增长导致的维度灾挑战，且由于该框架要求训练与执行阶段均集中式进行，在一些分布式场景下难以实际部署。

3) CTDE。

CTDE 框架通过将集中式训练与分布式执行结合，仅在训练中进行各智能体(部分)数据信息的共享以克服环境的非静态性，在完成训练后各智能体则依据本地观测进行自主决策和各智能体间的协同，无需估计各智能体的联合策略分布，从而避免了随着智能体数量增大导致的可扩展性问题。

由于 CTDE 框架同时解决了环境非静态性与智能体可扩展性问题，目前一些主流的 MADRL 算法，如 MADDPG、COMA、VDN 算法等均采用 CTDE 框架。同时，由于在执行阶段各智能体的分布式执行，可有效降低各决策单位的计算负担且利于在实际工程中部署。

此外，CTDE 框架又衍生出如下两种变种：

①参数共享(parameter sharing)^[88]：若各智能体具有相近的结构与策略优化目标(即同质智能体集群)，在集中训练阶段可以采用评价网络的参数共享机制以降低集中训练成本，提升算法的可扩展性。

②平均场近似(mean field approximation)：当智能体数量大幅增加时，由于维度的大幅度增加和智能体之间互动的复杂性增大，学习变得难以进行。文献[89]同样针对同质智能体集群，提出了基于平均场理论的 MARL 方法，对多智能体系统给出了一个近似假设：即对某个智能体，其他所有智能体对其产生的作用可以用一个均值替代。在此假设下，可以将一个智能体与其邻居智能体之间的相互作用简化为两个智能体之间的相互作用，极大地简化

了智能体数量带来的模型空间的增大。然而,该框架的应用前提是海量智能体间的通信连接拓扑已知,且各智能体可以获取与其相邻的智能体集合,并能与相邻智能体进行观测与动作等信息(包含产消者 DER 配置与运行状态、设备能量管理动作、能量交易与共享策略等个体隐私信息)的直接交互。在本文考虑的实际应用场景中,以上前置条件难以满足大规模产消者集群协同交易场景下所需具备的产消者隐私保护需求。

3.2.3 多智能体 DRL 算法在配用电系统中应用

文献[90-91]面向单个家庭与微网能量管理问题,构建各类型可调负荷的能量管理智能体,采用 MADRL 算法进行策略的协同,以实现家庭或微网的最优运行目标。MADRL 算法由于采用多个智能体相对独立的同时相互配合,更适用于配用电系统运行优化中能量交易、系统调控等多主体参与的决策场景。文献[92]依据微网中能量管理参与主体的不同分别设计了不同状态、动作、奖励,采用多智能体 DQN 算法进行训练,但各智能体不进行信息共享,在 DTDE 框架下难以避免环境的非静态性问题。文献[93]同样在 DTDE 框架下采用并行式学习方法对多智能体进行单独训练,由于各产消者仅关注自身的用电成本,其柔性负荷/发电的调控易集中在最低/最高电价时段,进而引发新的负荷/发电高峰,威胁配电网的安全稳定运行。

为克服 DTDE 框架弊端,文献[94]采用传统 CTDE 学习框架实现产消者间的分布式交易,通过在产消者之间进行全面的信息共享,消除了环境的非静态性,提升了策略的收敛性,但全面的信息共享难以满足产消者隐私保护的需求;同时,CTDE 框架下需要训练同产消者数量的神经网络,且各网络需要所有智能体的观测与动作信息作为输入,使得网络的规模随产消者数量剧增而引发维度灾。文献[95]采用 CTDE 结构下的多智能体 SAC 算法,分别构建各微网的连续动作决策智能体与离散动作决策智能体,实现了微网群分布式有、无功协同调度。文献[96]则直接将参数共享技术与 CTDE 学习框架融合,通过训练共享策略神经网络,实现了对大规模同质温控负荷的有效控制。文献[97]则针对多产消者参与的社区能量协同优化问题,进一步计及各产消者可控负荷、储能、电动汽车等多种可控资源的离散、连续混合调控动作,融合 CTDE 框架与参数共享机制,采用不同分布分别拟合离散、连

续动作分布提出多智能体优先深度确定性策略梯度算法(multi-agent prioritized deep deterministic policy gradient, MA-PDDPG),实现了产消者能量管理策略的协同优化,并有效促进了本地的光伏消纳与负荷尖峰抑制,增强了 MADRL 算法在社区能量管理问题中的适用性。

4 配用电系统运行优化挑战及展望

基于上述调研与分析可以得出,可交易能源市场环境配用电系统多层多主体复杂优化问题的研究方向,正由传统解析方法向 DRL 技术发展,但该技术电力市场及电力系统运行领域仍处于起步阶段,仍需要对如下几点关键问题开展深入研究:

1) 产消者能量管理方法难以适应复杂异构分布式资源的安全运行特性。

产消者的多元异构特性与异构分布式资源运行特性直接相关,并易受多种因素影响而呈现动态时变的多重内生不确定性。这些性质一方面增加了生成泛化性能量管理策略的难度,另一方面增加了策略违反用户设备安全运行约束的风险,难以保障其安全性。亟需具有泛化性策略自主学习能力的 DRL 技术予以辅助,实现考虑设备个体安全性的能量管理策略安全优化。

2) 多主体协同方法难以兼顾产消者异构性、隐私性、可扩展性及安全性。

在面对具有高维复杂特性的产消者集群协同优化运行问题时,现有多智能体强化学习方法存在为追求可扩展性而将产消者同质化处理的局限,同时缺乏良好的机制来避免产消者信息共享而导致的隐私泄露,更未考虑多产消者互动对个体安全约束的耦合影响。亟需在充分计及各产消者复杂异构性、隐私保护需求之上,设计可扩展的考虑个体安全性的多主体策略协同优化方法。

3) 配网调度管理方法难以满足市场化交易与调度运行的高效衔接需求。

可交易能源机制下,运营商面临多源状态观测、多元调控手段和全局服务目标的复杂决策特征,面临着市场主体与系统主体目标不一致的矛盾冲突。亟需设计相应的运营商-可交易能源市场的互动优化机制,充分考虑运营商观测及调控手段,提出市场与系统调和的配网运营管理方法。

为解决上述研究难题,本文提出了新一代深度

强化学习技术在可交易能源市场环境下配用电系统运行优化问题的研究思路，研究基于分层决策、自治协同的思想，针对贯穿于可交易能源市场环境下配用电系统安全运行优化问题中逐层加剧的多重不确定性、多主体协同需求与策略安全性需求等实际挑战开展针对性研究。基于深度强化学习策略生成方法，针对性地结合应对多重不确定性的多源时空信息感知方法、应对多主体协同需求的信息聚合与特征提取方法、应对策略安全性需求的策略安全性评价方法等应对方法与核心技术，将深度强化学习向多智能体、增强感知、有约束等方向进行改进，构建市场与系统异质主体策略的安全协同优化方法，如图 6 所示。所提出的市场环境下配用电系统中多层多主体安全协同优化方法有助于充分释放配网海量分布式资源的灵活调节能力，为解决可交易能源市场中大规模产消者策略的安全协同优化、市场交易与配用电系统安全运行高效衔接提供可行的技术实现路径。

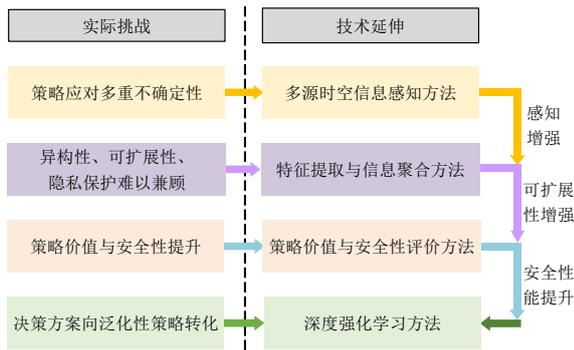


图 6 深度强化学习技术的延伸方案

Fig. 6 Envisaged extensions of DRL technologies

1) 个体产消者安全能量管理方法。

产消者的多元异构特性与异构分布式资源运行特性直接相关，并易受多种因素影响而呈现动态时变的多重内生不确定性。这些性质一方面增加了生成泛化性能量管理策略的难度，另一方面增加了策略违反用户设备安全运行约束的风险，难以保障其安全性。亟需具有泛化性策略自主学习能力的安安全深度强化学习技术予以辅助，实现考虑设备个体安全性的能量管理策略安全优化。

为解决传统决策模型面对产消者多元异构性与多重不确定性适用性不足的问题，针对多元异构产消者能量管理需求，首先需要构建产消者个体能量管理策略优化框架，将传统应用于无约束问题的 MDP 模型，改进为适用于产消者能量管理场景的带约束马尔科夫决策过程(constrained MDP, CMDP)

模型，通过额外引入约束评价函数、约束安全阈值来描述不同类型分布式资源的多种运行约束(储能系统充放电不可超过其总容量等设备运行特性约束，温控负荷所影响的室内温度需要处于一定舒适度范围等静态区间约束，电动汽车出行前电池荷电状态需要满足用户出行需求等时间耦合约束等)，从而进一步构建各运行约束在能量管理策略下的安全评价函数。其次，结合具有多源时序信息感知能力与高维混合连续动作决策能力的 DRL 智能体策略网络构建技术，结合泛用性分布式资源运行约束规则提取技术，构建计及安全约束的产消者个体能量管理策略优化方法，实现多元异构产消者的自适应能量管理。

2) 海量产消者交易策略可扩展安全协同方法。

在面对具有高维复杂特性的产消者集群协同优化运行问题时，现有多智能体强化学习方法存在为追求可扩展性而将产消者同质化处理的局限，同时缺乏良好的机制来避免产消者信息共享而导致的隐私泄露，更未考虑多产消者互动对个体安全约束的耦合影响。亟需在充分计及产消者复杂异质性、隐私保护需求之上，设计可扩展的考虑个体安全性的多主体策略协同优化方法。

为应对大规模能量交易市场模型中存在的多重外生不确定性耦合、用户隐私保护、可扩展性造成维度灾等挑战，需要基于产消者个体能量管理策略，实现从单产消者能量管理至产消者集群间的能量市场交易策略的扩展，产消者集群市场交易策略协同优化框架，将多智能体的 MG 模型改进为适用于多产消者能量交易场景的带约束马尔科夫博弈 (constrained MG, CMG)模型。

针对产消者本地信息共享导致的隐私泄露和产消者数量增多而导致的共享信息维度暴增的双重技术挑战，需要研究产消者对其本地隐私信息的加密技术，共享全局信息的降维技术，以及大规模智能体策略与价值网络的高效联合训练方法，最终实现个体产消者隐私保护和集群策略协同可扩展性的兼顾。最后，需要进一步研究多产消者互动行为安全规则提取技术，考虑多产消者互动对个体安全约束的耦合影响，提出相应的多智能体策略安全性评价方法，构建计及多产消者互动安全的可扩展性协同优化方法，实现大规模产消者间的能量管理与交易行为安全协同，促进配网中海量分布式资源的统筹优化。

3)可交易能源市场交易-配用电系统运行协同。

可交易能源机制下,运营商面临多源状态观测、多元调控手段和全局服务目标的复杂决策特征,面临着市场主体与系统主体目标不一致的矛盾冲突。亟需设计相应的运营商-可交易能源市场的互动优化机制,需要充分考虑运营商观测及调控手段,构建市场与系统调和的配网运营商管理方法。

针对市场交易与系统调度协同问题,需充分分析运营商与产消者集群的互动优化结构,基于CMDP模型构建配电网运营商与可交易能源市场交互协同优化框架。虽然现有研究对配用电系统中CMDP模型的构建应用已有初步研究^[98-101],但主要关注分布式资源接入场景下配电网电压、无功控制等问题,尚未能探索海量异构产消者配网市场分布式交易的复杂场景下的配电网调控问题。此外,需要分析可交易能源市场与配电网交织耦合的时空运行状态变化特性,构建具有时空耦合动态特性感知与高维混合动作能力的运营商智能体构建技术;并进一步分析潮流方程等配网安全运行约束,研究可交易能源市场环境下运营商调控策略的安全性评价方法,以构建促进市场化交易与配网安全运行高效衔接的运营商安全管理策略优化方法,实现可交易能源市场环境下配用电系统的安全管理。

5 结语

随着新型电力系统的不断推进,分布式灵活性资源在配电网的日益富集,可交易能源市场环境下配用电系统运行管理逐渐体发展为多层多主体高随机性复杂优化问题,传统的基于解析优化方法难以满足配用电系统运行管理需求。本文归纳总结了可交易能源市场环境下配用电系统运行调控中存在的挑战,对传统方法与人工智能方法在市场环境下配用电系统运行优化领域中的国内外研究现状进行了全面的综述,重点梳理并探讨了不同类型深度强化技术特点及其在研究领域内的适用性。最后,深入探讨了深度强化学习技术应用于市场环境下配用电系统的发展需求,展望了深度强化学习技术在实时性、可扩展性、隐私性、安全性方面发展的潜力与趋势。

参考文献

[1] 黄雨涵,丁涛,李雨婷,等.碳中和背景下能源低碳化技术综述及对新型电力系统发展的启示[J].中国电机工程学报,2021,41(S1):28-51.

HUANG Yuhang, DING Tao, LI Yuting, et al. Decarbonization technologies and inspirations for the development of novel power systems in the context of carbon neutrality[J]. Proceedings of the CSEE, 2021, 41(S1): 28-51(in Chinese).

[2] 国家能源局.国家能源局关于2021年风电、光伏发电开发建设有关事项的通知[EB/OL].北京:中国政府网,2021[2022-09-15].http://www.gov.cn/zhengce/zhengceku/2021-05/27/content_5612874.htm.

National Energy Administration. Notice of the National Energy Administration on matters related to the development and construction of wind power and photovoltaic power generation in 2021[EB/OL]. Beijing: GOV.CN, 2021[2022-09-15].http://www.gov.cn/zhengce/zhengceku/2021-05/27/content_5612874.htm(in Chinese).

[3] 刘念,余星火,王剑辉,等.泛在物联网的配用电优化运行:信息物理社会系统的视角[J].电力系统自动化,2020,44(1):1-12.

LIU Nian, YU Xinghuo, WANG Jianhui, et al. Optimal operation of power distribution and consumption system based on ubiquitous internet of things: a cyber-physical-social system perspective[J]. Automation of Electric Power Systems, 2020, 44(1): 1-12(in Chinese).

[4] 国家发展改革委.国家能源局关于加快建设全国统一电力市场体系的指导意见[EB/OL].北京:中国政府网,2022[2022-09-15].http://www.gov.cn/zhengce/zhengceku/2022-01/30/content_5671296.htm.

National Development and Reform Commission. Guidance from the national energy administration on accelerating the construction of a national unified electricity market system [EB/OL]. Beijing: GOV.CN, 2022 [2022-09-15].http://www.gov.cn/zhengce/zhengceku/2022-01/30/content_5671296.htm(in Chinese).

[5] 国家能源局.国家能源局2022年深化“放管服”改革优化营商环境重点任务分工方案[EB/OL].北京:中国政府网,2022[2022-09-15].http://www.nea.gov.cn/2022-08/25/c_1310655795.htm.

National Energy Administration. The National Energy Administration's key task division plan for deepening the “deregulation, regulation and service” reform and optimizing the business environment in 2022 [EB/OL]. Beijing: GOV.CN, 2022[2022-09-15].http://www.nea.gov.cn/2022-08/25/c_1310655795.htm(in Chinese).

[6] 万灿,贾妍博,李彪,等.城镇能源互联网能源交易模式和用户响应研究现状与展望[J].电力系统自动化,2019,43(14):29-40.

WAN Can, JIA Yanbo, LI Biao, et al. Research status and

- prospect of energy trading mode and user demand response in urban energy internet[J]. *Automation of Electric Power Systems*, 2019, 43(14): 29-40(in Chinese).
- [7] 陈启鑫, 王克道, 陈思捷, 等. 面向分布式主体的可交易能源系统: 体系架构、机制设计与关键技术[J]. *电力系统自动化*, 2018, 42(3): 1-7, 31.
CHEN Qixin, WANG Kedao, CHEN Sijie, et al. Transactive energy system for distributed agents: architecture, mechanism design and key technologies [J]. *Automation of Electric Power Systems*, 2018, 42(3): 1-7, 31(in Chinese).
- [8] 胡俊杰, 王坤宇, 艾欣, 等. 交互能源: 实现电力能源系统平衡的有效机制[J]. *中国电机工程学报*, 2019, 39(4): 953-965.
HU Junjie, WANG Kunyu, AI Xin, et al. Transactive energy: an effective mechanism for balancing electric energy system[J]. *Proceedings of the CSEE*, 2019, 39(4): 953-965(in Chinese).
- [9] 陈皓勇, 谭碧飞, 伍亮, 等. 分层集群的新型电力系统运行与控制[J]. *中国电机工程学报*, 2023, 43(2): 581-595.
CHEN Haoyong, TAN Bifei, WU Liang, et al. Operation and control of the new power systems based on hierarchical clusters[J]. *Proceedings of the CSEE*, 2023, 43(2): 581-595(in Chinese).
- [10] Special Section on local and distribution electricity markets[EB/OL]. *IEEE Transactions on Smart Grid*, 2021[2022-09-15]. https://www.ieee-pes.org/images/files/TSG_CFP_Local_and_Distribution_Electricity_Market_1.pdf.
- [11] YUEN C, SAHA T K, POOR H V, et al. Energy sharing in the peer-to-peer network-moving towards real implementation[EB/OL]. *Applied Energy*, 2021[2022-09-15]. <https://www.sciencedirect.com/journal/applied-energy/special-issue/10HVGXMQQLZQ>.
- [12] 中国电机工程学报. 面向新型电力系统的配电市场发展路径及关键技术专题[EB/OL]. *中国电机工程学报*, 2022[2022-09-15]. http://www.pcsee.org/YUdz_d8A_tm8HcJZSQfdypak1eI93Mf%2BootcVSRgQe8%3D?encrypt=1.
Proceedings of the CSEE. Special section on development path and key technologies of distribution market for new power systems[EB/OL]. *Proceedings of the CSEE*, 2022[2022-09-15]. http://www.pcsee.org/YUdz_d8A_tm8HcJZSQfdypak1eI93Mf%2BootcVSRgQe8%3D?encrypt=1(in Chinese).
- [13] 电力系统自动化. 分布式源荷参与辅助服务专题专辑
征稿启事[EB/OL]. *电力系统自动化*, 2022[2022-09-15]. <http://www.aeps-info.com/aeps/news/view/20220901091228001>.
Automation of Electric Power Systems. Special section on distributed source and load participating ancillary services[EB/OL]. *Automation of Electric Power Systems*, 2022[2022-09-15]. <http://www.aeps-info.com/aeps/news/view/20220901091228001>(in Chinese).
- [14] PATERAKIS N G, ERDINÇ O, BAKIRTZIS A G, et al. Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies[J]. *IEEE Transactions on Industrial Informatics*, 2015, 11(6): 1509-1519.
- [15] ANVARI-MOGHADDAM A, MONSEF H, RAHIMI-KIAN A. Optimal smart home energy management considering energy saving and a comfortable lifestyle [J]. *IEEE Transactions on Smart Grid*, 2015, 6(1): 324-332.
- [16] 陈中, 李云倩, 冷钊莹, 等. 典型家用大功率负载精细化建模及能量管理策略[J]. *电力系统自动化*, 2018, 42(22): 135-143.
CHEN Zhong, LI Yunqian, LENG Zhaoying, et al. Refined modeling and energy management strategy of typical household high-power loads[J]. *Automation of Electric Power Systems*, 2018, 42(22): 135-143(in Chinese).
- [17] 陆青, 郁浩, 冷亚军, 等. 家庭智能用电任务调度优化模型及其算法研究[J]. *中国电机工程学报*, 2018, 38(13): 3826-3836, 4023.
LU Qing, YU Hao, LENG Yajun, et al. Research on model and algorithm of smart electricity consumption task scheduling optimization in household[J]. *Proceedings of the CSEE*, 2018, 38(13): 3826-3836, 4023(in Chinese).
- [18] 孙毅, 李泽坤, 许鹏, 等. 异构柔性负荷建模调控关键技术及发展方向研究[J]. *中国电机工程学报*, 2019, 39(24): 7146-7158, 7488.
SUN Yi, LI Zekun, XU Peng, et al. Research on key technologies and development direction of heterogeneous flexible load modeling and regulation[J]. *Proceedings of the CSEE*, 2019, 39(24): 7146-7158, 7488(in Chinese).
- [19] 吴界辰, 艾欣. 交互能源机制下的电力产消者优化运行[J]. *电力系统自动化*, 2020, 44(19): 1-18.
WU Jiechen, AI Xin. Optimal operation of prosumers based on transactive energy mechanism[J]. *Automation of Electric Power Systems*, 2020, 44(19): 1-18(in Chinese).
- [20] LI Jie, LIU Yikui, WU Lei. Optimal operation for

- community-based multi-party microgrid in grid-connected and islanded modes[J]. *IEEE Transactions on Smart Grid*, 2018, 9(2): 756-765.
- [21] CHEN Zhi, WU Lei, FU Yong. Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization [J]. *IEEE Transactions on Smart Grid*, 2012, 3(4): 1822-1831.
- [22] SU Wencong, WANG Jianhui, ROH J. Stochastic energy scheduling in microgrids with intermittent renewable energy resources[J]. *IEEE Transactions on Smart Grid*, 2014, 5(4): 1876-1883.
- [23] SHAFIE-KHAH M, SIANO P. A stochastic home energy management system considering satisfaction cost and response fatigue[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(2): 629-638.
- [24] DU Y F, JIANG Lin, LI Yuanzheng, et al. A robust optimization approach for demand side scheduling considering uncertainty of manually operated appliances [J]. *IEEE Transactions on Smart Grid*, 2018, 9(2): 743-755.
- [25] 于丹文, 杨明, 翟鹤峰, 等. 鲁棒优化在电力系统调度决策中的应用研究综述[J]. *电力系统自动化*, 2016, 40(7): 134-143, 148.
- YU Danwen, YANG Ming, ZHAI Hefeng, et al. An overview of robust optimization used for power system dispatch and decision-making[J]. *Automation of Electric Power Systems*, 2016, 40(7): 134-143, 148(in Chinese).
- [26] 魏韡, 刘锋, 梅生伟. 电力系统鲁棒经济调度(一)理论基础[J]. *电力系统自动化*, 2013, 37(17): 37-43.
- WEI Wei, LIU Feng, MEI Shengwei. Robust and economical scheduling methodology for power systems part one theoretical foundations[J]. *Automation of Electric Power Systems*, 2013, 35(17): 37-43(in Chinese).
- [27] NING Chao, YOU Fengqi. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming [J]. *Computers & Chemical Engineering*, 2019, 125: 434-448.
- [28] TUSHAR W, YUEN C, SAHA T K, et al. Peer-to-peer energy systems for connected communities: A review of recent advances and emerging challenges[J]. *Applied Energy*, 2021, 282: 116131.
- [29] 李山山, 李华强, 金智博, 等. 基于共享经济理念的园区分布式能源共享服务机制[J]. *中国电机工程学报*, 2022, 42(1): 56-70.
- LI Shanshan, LI Huaqiang, JIN Zhibo, et al. Distributed energy sharing service mechanism for park based on the concept of sharing economy[J]. *Proceedings of the CSEE*, 2022, 42(1): 56-70(in Chinese).
- [30] ALAM M R, ST-HILAIRE M, KUNZ T. Peer-to-peer energy trading among smart homes[J]. *Applied Energy*, 2019, 238: 1434-1443.
- [31] ZHOU Suyang, ZOU Fenghua, WU Zhi, et al. A smart community energy management scheme considering user dominated demand side response and P2P trading [J]. *International Journal of Electrical Power & Energy Systems*, 2020, 114: 105378.
- [32] LEZAMA F, SOARES J, HERNANDEZ-LEAL P, et al. Local energy markets: paving the path toward fully transactive energy systems[J]. *IEEE Transactions on Power Systems*, 2019, 34(5): 4081-4088.
- [33] ONUGHA I, DEGHAN S, ARISTIDOU P. Rethinking consumer-centric markets under uncertainty: A robust approach to community-based energy trades[C]//2020 IEEE Power & Energy Society General Meeting (PESGM). Montreal, QC, Canada: IEEE, 2020: 1-5.
- [34] LI Jing, YE Yujian, PAPADASKALOPOULOS D, et al. Distributed consensus-based coordination of flexible demand and energy storage resources[J]. *IEEE Transactions on Power Systems*, 2021, 36(4): 3053-3069.
- [35] SORIN E, BOBO L, PINSON P. Consensus-based approach to peer-to-peer electricity markets with product differentiation[J]. *IEEE Transactions on Power Systems*, 2019, 34(2): 994-1004.
- [36] MORSTYN T, TEYTELBOYM A, MCCULLOCH M D. Bilateral contract networks for peer-to-peer energy trading[J]. *IEEE Transactions on Smart Grid*, 2019, 10(2): 2026-2035.
- [37] CHAKRABORTY S, BAARSLAG T, KAISERS M. Automated peer-to-peer negotiation for energy contract settlements in residential cooperatives[J]. *Applied Energy*, 2020, 259: 114173.
- [38] 张虹, 闫贺, 申鑫, 等. 面向能源社区能量管理的配网产消者分布式优化调度[J]. *中国电机工程学报*, 2022, 42(12): 4449-4458.
- ZHANG Hong, YAN He, SHEN Xin, et al. Distributed optimal scheduling for prosumer in distribution network for energy community energy management [J]. *Proceedings of the CSEE*, 2022, 42(12): 4449-4458(in Chinese).
- [39] CRESPO-VAZQUEZ J L, ALSKAIF T, GONZÁLEZ-RUEDA Á M, et al. A community-based energy market design using decentralized decision-making under uncertainty[J]. *IEEE Transactions on Smart Grid*, 2021, 12(2): 1782-1793.

- [40] 高红均, 张凡, 刘俊勇, 等. 考虑多产消者差异化特征的社区微网系统 P2P 交易机制设计[J]. 中国电机工程学报, 2022, 42(4): 1455-1469.
GAO Hongjun, ZHANG Fan, LIU Junyong, et al. Design of P2P transaction mechanism considering differentiation characteristics of multiple prosumers in community microgrid system[J]. Proceedings of the CSEE, 2022, 42(4): 1455-1469(in Chinese).
- [41] 王健, 周念成, 王强钢, 等. 基于区块链和连续双向拍卖机制的微电网直接交易模式及策略[J]. 中国电机工程学报, 2018, 38(17): 5072-5084.
WANG Jian, ZHOU Niancheng, WANG Qianggang, et al. Electricity direct transaction mode and strategy in microgrid based on Blockchain and continuous double auction mechanism[J]. Proceedings of the CSEE, 2018, 38(17): 5072-5084(in Chinese).
- [42] DENG Ruilong, YANG Zaiyue, HOU Fen, et al. Distributed real-time demand response in multiseller-multibuyer smart distribution grid[J]. IEEE Transactions on Power Systems, 2015, 30(5): 2364-2374.
- [43] WANG Zhenyu, ZHANG Xiaoyu, WANG Hao. Consensus-based decentralized energy trading for distributed energy resources[J]. Energy Conversion and Economics, 2021, 2(4): 221-234.
- [44] MORSTYN T, MCCULLOCH M D. Multiclass energy management for peer-to-peer energy trading driven by prosumer preferences[J]. IEEE Transactions on Power Systems, 2019, 34(5): 4005-4014.
- [45] KHORASANY M, NAJAFI-GHALELOU A, RAZZAGHI R. A framework for joint scheduling and power trading of prosumers in transactive markets [J]. IEEE Transactions on Sustainable Energy, 2021, 12(2): 955-965.
- [46] MORSTYN T, TEYTELBOYM A, HEPBURN C, et al. Integrating P2P energy trading with probabilistic distribution locational marginal pricing[J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3095-3106.
- [47] 陈修鹏, 李庚银, 夏勇. 基于主从博弈的新型城镇配电网产消者竞价策略[J]. 电力系统自动化, 2019, 43(14): 97-104.
CHEN Xiupeng, LI Gengyin, XIA Yong. Stackelberg game based bidding strategy for prosumers in new urban distribution system[J]. Automation of Electric Power Systems, 2019, 43(14): 97-104(in Chinese).
- [48] 叶宇剑, 袁泉, 汤奕, 等. 抑制柔性负荷过响应的微网分散式调控参数优化[J]. 中国电机工程学报, 2022, 42(5): 1748-1759.
YE Yujian, YUAN Quan, TANG Yi, et al. Decentralized coordination parameters optimization in microgrids mitigating demand response synchronization effect of flexible loads[J]. Proceedings of the CSEE, 2022, 42(5): 1748-1759(in Chinese).
- [49] AZIM M I, TUSHAR W, SAHA T K. Coalition graph game-based P2P energy trading with local voltage management[J]. IEEE Transactions on Smart Grid, 2021, 12(5): 4389-4402.
- [50] QIU Dawei, YE Yujian, PAPADASKALOPOULOS D. Exploring the effects of local energy markets on electricity retailers and customers[J]. Electric Power Systems Research, 2020, 189: 106761.
- [51] WANG Yifei, HUANG Zhiheng, SHAHIDEHPOUR M, et al. Reconfigurable distribution network for managing transactive energy in a multi-microgrid system[J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1286-1295.
- [52] GRIMM V, ORLINSKAYA G, SCHEWE L, et al. Optimal design of retailer-prosumer electricity tariffs using bilevel optimization[J]. Omega, 2021, 102: 102327.
- [53] SEKIZAKI S, NISHIZAKI I, HAYASHIDA T. Electricity retail market model with flexible price settings and elastic price-based demand responses by consumers in distribution network[J]. International Journal of Electrical Power & Energy Systems, 2016, 81: 371-386.
- [54] LIU Wenjie, CHEN Shibo, HOU Yunhe, et al. Optimal reserve management of electric vehicle aggregator: Discrete bilevel optimization model and exact algorithm [J]. IEEE Transactions on Smart Grid, 2021, 12(5): 4003-4015.
- [55] LIU Nian, YU Xinghuo, WANG Cheng, et al. Energy sharing management for microgrids with PV prosumers: A Stackelberg game approach[J]. IEEE Transactions on Industrial Informatics, 2017, 13(3): 1088-1098.
- [56] ZHANG Yuanyuan, ZHAO Huiru, LI Bingkang, et al. Research on dynamic pricing and operation optimization strategy of integrated energy system based on Stackelberg game[J]. International Journal of Electrical Power & Energy Systems, 2022, 143: 108446.
- [57] VINYALS O, BABUSCHKIN I, CZARNECKI W M, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning[J]. Nature, 2019, 575(7782): 350-354.
- [58] 赵晋泉, 夏雪, 徐春雷, 等. 新一代人工智能技术在电力系统调度运行中的应用评述[J]. 电力系统自动化, 2020, 44(24): 1-10.
ZHAO Jinquan, XIA Xue, XU Chunlei, et al. Review on

- application of new generation artificial intelligence technology in power system dispatching and operation [J]. Automation of Electric Power Systems, 2020, 44(24): 1-10(in Chinese).
- [59] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. 2nd ed. Cambridge, MA: MIT Press, 2018.
- [60] 汤涌, 姚伟, 王宏志, 等. 电网仿真分析与决策的人工智能方法[J]. 中国电机工程学报, 2022, 42(15): 5384-5405.
- TANG Yong, YAO Wei, WANG Hongzhi, et al. Artificial intelligence techniques for power grid simulation analysis and decision making[J]. Proceedings of the CSEE, 2022, 42(15): 5384-5405(in Chinese).
- [61] 尚宇炜, 郭剑波, 吴文传, 等. 电力脑初探: 一种多模态自适应学习系统[J]. 中国电机工程学报, 2018, 38(11): 3133-3143.
- SHANG Yuwei, GUO Jianbo, WU Wenchuan, et al. Preliminary study of electric power brain: a multimodal adaptive learning system[J]. Proceedings of the CSEE, 2018, 38(11): 3133-3143(in Chinese).
- [62] 乔骥, 王新迎, 张擎, 等. 基于柔性行动器-评判器深度强化学习的电气综合能源系统优化调度[J]. 中国电机工程学报, 2021, 41(3): 819-832.
- QIAO Ji, WANG Xinying, ZHANG Qing, et al. Optimal dispatch of integrated electricity-gas system with soft actor-critic deep reinforcement learning[J]. Proceedings of the CSEE, 2021, 41(3): 819-832(in Chinese).
- [63] 彭刘阳, 孙元章, 徐箭, 等. 基于深度强化学习的自适应不确定性经济调度[J]. 电力系统自动化, 2020, 44(9): 33-42.
- PENG Liuyang, SUN Yuanzhang, XU Jian, et al. Self-adaptive uncertainty economic dispatch based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2020, 44(9): 33-42(in Chinese).
- [64] 李琦, 乔颖, 张宇精. 配电网持续无功优化的深度强化学习方法[J]. 电网技术, 2020, 44(4): 1473-1480.
- LI Qi, QIAO Ying, ZHANG Yujing. Continuous reactive power optimization of distribution network using deep reinforcement learning[J]. Power System Technology, 2020, 44(4): 1473-1480(in Chinese).
- [65] ZHANG Ying, WANG Xinan, WANG Jianhui, et al. Deep reinforcement learning based volt-VAR optimization in smart distribution systems[J]. IEEE Transactions on Smart Grid, 2021, 12(1): 361-371.
- [66] 倪爽, 崔承刚, 杨宁, 等. 基于深度强化学习的配电网多时间尺度在线无功优化[J]. 电力系统自动化, 2021, 45(10): 77-85.
- NI Shuang, CUI Chenggang, YANG Ning, et al. Multi-time-scale online optimization for reactive power of distribution network based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2021, 45(10): 77-85(in Chinese).
- [67] WANG Biao, LI Yan, MING Weiyu, et al. Deep reinforcement learning method for demand response management of interruptible load[J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3146-3155.
- [68] 龚锦霞, 刘艳敏. 基于深度确定策略梯度算法的主动配电网协调优化[J]. 电力系统自动化, 2020, 44(6): 113-120.
- GONG Jinxia, LIU Yanmin. Deterministic policy gradient algorithm coordinated optimization of active distribution network based on deep[J]. Automation of Electric Power Systems, 2020, 44(6): 113-120(in Chinese).
- [69] 黄玉雄, 李更丰, 张理寅, 等. 弹性配电系统动态负荷恢复的深度强化学习方法[J]. 电力系统自动化, 2022, 46(8): 68-78.
- HUANG Yuxiong, LI Gengfeng, ZHANG Liyin, et al. Deep reinforcement learning method for dynamic load restoration of resilient distribution systems[J]. Automation of Electric Power Systems, 2022, 46(8): 68-78(in Chinese).
- [70] ZHAO Tianqiao, WANG Jianhui. Learning sequential distribution system restoration via graph-reinforcement learning[J]. IEEE Transactions on Power Systems, 2022, 37(2): 1601-1611.
- [71] ZHOU Zhenchen, WU Zhou, JIN Tao. Deep reinforcement learning framework for resilience enhancement of distribution systems under extreme weather events[J]. International Journal of Electrical Power & Energy Systems, 2021, 128: 106676.
- [72] 王子昊, 王旭, 蒋传文, 等. 基于近端策略优化算法的灾后配电网韧性提升方法[J]. 电力系统自动化, 2022, 46(21): 62-70.
- WANG Zihao, WANG Xu, JIANG Chuanwen, et al. Resilience improvement method for post-disaster distribution network based on proximal policy optimization algorithm[J]. Automation of Electric Power Systems, 2022, 46(21): 62-70(in Chinese).
- [73] JI Ying, WANG Jianhui, XU Jiaca, et al. Real-time energy management of a microgrid using deep reinforcement learning[J]. Energies, 2019, 12(12): 2291.
- [74] 刘俊峰, 陈剑龙, 王晓生, 等. 基于深度强化学习的微能源网能量管理与优化策略研究[J]. 电网技术, 2020,

- 44(10): 3794-3803.
- LIU Junfeng, CHEN Jianlong, WANG Xiaosheng, et al. Energy management and optimization of multi-energy grid based on deep reinforcement learning[J]. Power System Technology, 2020, 44(10): 3794-3803(in Chinese).
- [75] 黎海涛, 申保晨, 杨艳红, 等. 基于改进竞争深度 Q 网络算法的微电网能量管理与优化策略[J]. 电力系统自动化, 2022, 46(7): 42-49.
- LI Haitao, SHEN Baochen, YANG Yanhong, et al. Energy management and optimization strategy for microgrid based on improved dueling deep Q network algorithm [J]. Automation of Electric Power Systems, 2022, 46(7): 42-49(in Chinese).
- [76] 张甜, 赵奇, 陈中, 等. 基于深度强化学习的家庭能量管理分层优化策略[J]. 电力系统自动化, 2021, 45(21): 149-158.
- ZHANG Tian, ZHAO Qi, CHEN Zhong, et al. Hierarchical optimization strategy for home energy management based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2021, 45(21): 149-158(in Chinese).
- [77] YU Liang, XIE Weiwei, XIE Di, et al. Deep reinforcement learning for smart home energy management[J]. IEEE Internet of Things Journal, 2020, 7(4): 2751-2762.
- [78] ZHANG Guozhou, HU Weihao, CAO Di, et al. Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach [J]. Energy Conversion and Management, 2021, 227: 113608.
- [79] 叶宇剑, 王卉宇, 汤奕, 等. 基于深度强化学习的居民实时自治最优能量管理策略[J]. 电力系统自动化, 2022, 46(1): 110-119.
- YE Yujian, WANG Huiyu, TANG Yi, et al. Real-time autonomous optimal energy management strategy for residents based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2022, 46(1): 110-119(in Chinese).
- [80] LU Renzhi, HONG S H, ZHANG Xiongfeng. A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach[J]. Applied Energy, 2018, 220: 220-230.
- [81] ZHANG Qianzhi, DEGHANPOUR K, WANG Zhaoyu, et al. A learning-based power management method for networked microgrids under incomplete information [J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1193-1204.
- [82] XU Hanchen, SUN Hongbo, NIKOVSKI D, et al. Deep reinforcement learning for joint bidding and pricing of load serving entity[J]. IEEE Transactions on Smart Grid, 2019, 10(6): 6366-6375.
- [83] QIU Dawei, YE Yujian, PAPADASKALOPOULOS D, et al. A deep reinforcement learning method for pricing electric vehicles with discrete charging levels[J]. IEEE Transactions on Industry Applications, 2020, 56(5): 5901-5912.
- [84] WANG Shengyi, DUAN Jiajun, SHI Di, et al. A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning[J]. IEEE Transactions on Power Systems, 2020, 35(6): 4644-4654.
- [85] CAO Di, ZHAO Junbo, HU Weihao, et al. Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs[J]. IEEE Transactions on Smart Grid, 2021, 12(5): 4137-4150.
- [86] 姚明涛, 胡兆光, 张宁, 等. 工业负荷提供辅助服务的多智能体响应模拟[J]. 中国电机工程学报, 2014, 34(25): 4219-4226.
- YAO Mingtao, HU Zhaoguang, ZHANG Ning, et al. Multi-agent response simulation of industrial loads for ancillary services[J]. Proceedings of the CSEE, 2014, 34(25): 4219-4226(in Chinese).
- [87] 田建伟, 胡兆光, 吴俊勇, 等. 基于多智能体建模的经济-电力动态模拟系统[J]. 中国电机工程学报, 2010, 30(7): 85-91.
- TIAN Jianwei, HU Zhaoguang, WU Junyong, et al. Dynamic economy and power simulation system based on multi-agent modelling[J]. Proceedings of the CSEE, 2010, 30(7): 85-91(in Chinese).
- [88] GUPTA J K, EGOV M, KOCHENDERFER M. Cooperative multi-agent control using deep reinforcement learning[C]//AAMAS 2017 Workshops on Autonomous Agents and Multiagent Systems. São Paulo, Brazil: Springer, 2017.
- [89] YANG Yaodong, LUO Rui, LI Minne, et al. Mean field multi-agent reinforcement learning[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018: 5571-5580.
- [90] KOFINAS P, DOUNIS A J, VOUIROS G A. Fuzzy Q-learning for multi-agent decentralized energy management in microgrids[J]. Applied Energy, 2018, 219: 53-67.
- [91] XU Xu, JIA Youwei, XU Yan, et al. A multi-agent reinforcement learning-based data-driven method for

- home energy management[J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3201-3211.
- [92] FANG Xiaohan, ZHAO Qiang, WANG Jinkuan, et al. Multi-agent deep reinforcement learning for distributed energy management and strategy optimization of microgrid market[J]. Sustainable Cities and Society, 2021, 74: 103163.
- [93] ZHANG Chi, KUPPANNAGARI S R, XIONG Chuanxiu, et al. A cooperative multi-agent deep reinforcement learning framework for real-time residential load scheduling[C]//Proceedings of the International Conference on Internet of Things Design and Implementation. New York, NY, USA: Association for Computing Machinery, 2019: 59-69.
- [94] CHUNG H M, MAHARJAN S, ZHANG Yan, et al. Distributed deep reinforcement learning for intelligent load scheduling in residential smart grids[J]. IEEE Transactions on Industrial Informatics, 2021, 17(4): 2752-2763.
- [95] 巨云涛, 陈希. 基于双层多智能体强化学习的微网群分布式有功无功协调优化调度[J]. 中国电机工程学报, 2022, 42(23): 8534-8548.
- JU Yuntao, CHEN Xi. Distributed active and reactive power coordinated optimal scheduling of networked microgrids based on two-layer multi-agent reinforcement learning[J]. Proceedings of the CSEE, 2022, 42(23): 8534-8548(in Chinese).
- [96] KAZMI H, SUYKENS J, BALINT A, et al. Multi-agent reinforcement learning for modeling and control of thermostatically controlled loads[J]. Applied Energy, 2019, 238: 1022-1035.
- [97] 叶宇剑, 袁泉, 刘文雯, 等. 基于参数共享机制多智能体深度强化学习的社区能量管理协同优化[J]. 中国电机工程学报, 2022, 42(21): 7682-7694.
- YE Yujian, YUAN Quan, LIU Wenwen, et al. Parameter sharing empowered multi-agent deep reinforcement learning for coordinated management of energy communities[J]. Proceedings of the CSEE, 2022, 42(21): 7682-7694(in Chinese).
- [98] LI Hepeng, HE Haibo. Learning to operate distribution networks with safe deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2022, 13(3): 1860-1872.
- [99] LI Hepeng, WAN Zhiqiang, HE Haibo. Constrained EV charging scheduling based on safe deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2427-2439.
- [100] GAO Yuanqi, YU Nanpeng. Model-augmented safe reinforcement learning for Volt-VAR control in power distribution networks[J]. Applied Energy, 2022, 313: 118762.
- [101] YE Yujian, WANG Hongru, CHEN Peiling, et al. Safe deep reinforcement learning for microgrid energy management in distribution networks with leveraged spatial-temporal perception[J/OL]. IEEE Transactions on Smart Grid, 2023; 1-1[2023-02-07]. <https://ieeexplore.ieee.org/abstract/document/10040615>.



叶宇剑

在线出版日期: 2023-10-10。

收稿日期: 2023-02-21。

作者简介:

叶宇剑(1988), 男, 教授, 博士生导师, 研究方向为电力市场建模与分析, 人工智能在新型电力系统的应用等, yeyujian@seu.edu.cn;

吴奕之(2000), 男, 硕士研究生, 研究方向为配电网优化运行与韧性提升, 220222674@seu.edu.cn;

胡健雄(1996), 男, 博士研究生, 主要研究方向为电力系统稳定分析与控制、人工智能在电力系统中的应用, 220182640@seu.edu.cn;

汤奕(1977), 男, 教授, 博士生导师, 研究方向为电力系统稳定分析、电力信息物理系统等, tangyi@seu.edu.cn;

陈涛(1989), 男, 副研究员, 硕士生导师, 研究方向为配网侧需求管理, taoc@seu.edu.cn;

Goran Strbac(1959), 男, 教授, 博士生导师, 研究方向为能源系统运行与规划等, g.strbac@imperial.ac.uk。

(责任编辑 李泽荣)

Hierarchical Coordinated Optimization for Power Distribution and Consumption System Operation in a Market Environment: Challenges, Progress and Prospects

YE Yujian¹, WU Yizhi¹, HU Jianxiong¹, TANG Yi¹, CHEN Tao¹, Goran STRBAC²

(1. School of Electrical Engineering, Southeast University;

2. Department of Electrical and Electronic Engineering, Imperial College London)

KEY WORDS: distribution market operation; distribution system dispatch; transactive energy; demand side management; reinforcement learning; multi-agent systems

With the increasing proliferation of distributed energy resources in the distribution network, how to establish an effective market-based trading mechanism in the power distribution and consumption system, while achieving efficient and coordinated optimization of market trading and power system operation has attracted unprecedented research interests in China and beyond. Under the market environment, the operation and management of each layer of the power distribution and consumption system (Fig. 1) face multiple challenges, including the layer-wise increasing uncertainties, the increasing scale and volume of market transactions, and lack of efficient coordination of market trading and safe operation of the system.

In order to address these challenges, this paper firstly outlines the critical scientific problems associated with optimal operation of power distribution and

consumption systems in a market environment. Second, it critically reviews and summarizes existing research efforts in this area, employing conventional, mathematical-optimization-based solution techniques, and subsequently concludes remaining issues that deserve further research attention.

Going further, this paper comprehensively reviews relevant deep reinforcement learning techniques and outlines their current applications in the examined research area, considering the primary characteristics pertaining to market trading and dispatch challenges associated with distribution and consumption system.

Finally, this paper details three directions which requires further research efforts, and also dives deep in revealing how deep reinforcement learning techniques can be developed and extended to support relevant research activities.

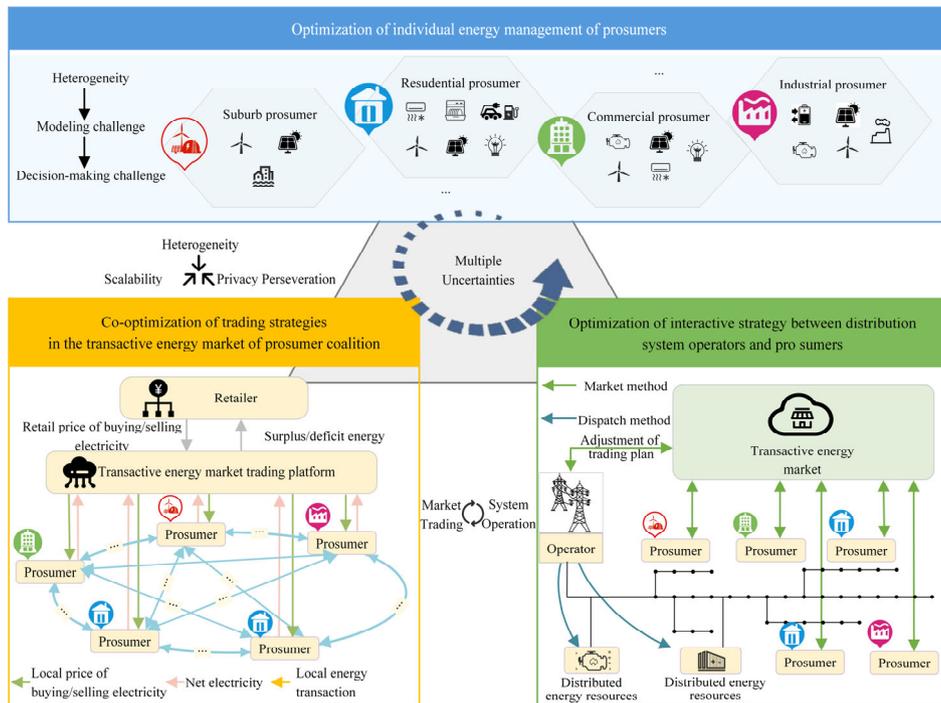


Fig. 1 Optimization of power distribution and consumption system operation in a transactive energy market environment